JOURNAL OF APPLIED BEHAVIOR ANALYSIS 1999, **32,** 25–33 NUMBER 1 (SPRING 1999)

REDUCTION OF STIMULUS OVERSELECTIVITY WITH NONVERBAL DIFFERENTIAL OBSERVING RESPONSES

WILLIAM V. DUBE AND WILLIAM J. MCILVANE

E. K. SHRIVER CENTER FOR MENTAL RETARDATION
AND NORTHEASTERN UNIVERSITY

Three individuals with mental retardation exhibited stimulus overselectivity in a delayed matching-to-sample task in which two sample stimuli were displayed on each trial. Intermediate accuracy scores indicated that participants could match one of the samples but not both of them. Accuracy in a baseline condition was compared to accuracy with a differential observing response procedure. This procedure prompted participants to make simultaneous identity-matching responses that required observation and discrimination of both sample stimuli. These observing responses were never followed by differential consequences. When observing responses were prompted, participants' accuracy scores improved. In a return to the baseline condition, when differential observing responses were no longer prompted, accuracy returned to intermediate levels. The results show that stimulus overselectivity can be greatly reduced by a behavioral intervention that controls observing behavior and verifies discrimination, but that exposure to such procedures alone may be insufficient for lasting benefits.

DESCRIPTORS: stimulus overselectivity, differential observing responses, matching to sample, mentally retarded

This paper reports a bridge study (Wacker, 1996), a basic analogue experiment that addresses a socially relevant problem known as *stimulus overselectivity* or *restricted stimulus control.* Overselectivity refers to atypically limited learning with respect to range, breadth, or number of stimuli or stimulus features (Lovaas, Koegel, & Schreibman, 1979). It is shown when training with multiple stimuli results in stimulus control by an atypically limited subset of those stimuli. Overselectivity is a widely acknowledged problem in the education of individuals with developmental disabilities like mental retardation and autism (e.g., Allen & Fuqua, 1985; Bickel, Richmond, Bell, & Brown, 1986). Although the problem is widespread, the remediation literature is remarkably small (for a recent summary, see Schreibman, 1997).

Overselectivity may affect performances on matching-to-sample tasks in special education classrooms. Matching to sample is widely used in such settings to teach stimulus–stimulus relations among spoken and printed words, objects, pictures, and sometimes symbols used in augmentative and alternative communication systems (e.g., de Rose, de Souza, & Hanna, 1996; Stromer, Mackay, & Stoddard, 1992). For example, consider a student who is learning to identify her printed name. If discriminative control by the name were restricted to the initial letter only, the student may discriminate SUE from FAY or BOB, but she would fail to discriminate SUE from SAM reliably (cf. Birnie-Selwyn & Guerin, 1997).

For several years, our laboratory has been studying the problem of overselectivity using

Data collection and manuscript preparation were supported by NICHD Grant HD 25995. For their help with data collection or analysis, we thank Lyn Balsamo, Kathy Clark, Kevin Farren, Tom Fowler, Kristin Lombard, Alison McVay, Jennifer Mlocek, Nora Murphy, and Aimee Smith. For their cooperation, we also thank The Protestant Guild Learning Center, The New England Center for Children, and The Behavior Intervention Project.

Address correspondence to William V. Dube, Psychological Sciences Division, E. K. Shriver Center, 200 Trapelo Road, Waltham, Massachusetts 02254 (E-mail: wdube@shriver.org).

an analogue task, delayed matching to sample (DMTS) with multiple sample stimuli (e.g., Stromer, McIlvane, Dube, & Mackay, 1993). On each trial, two sample stimuli are presented (e.g., AB). The samples remain available for observation until the participant touches them; then they disappear and the comparison stimuli are presented immediately (i.e., with a 0-s delay between sample offset and comparison onset). The comparisons are three individual stimuli, one of which is identical to one of the sample stimuli (e.g., A, C, and D). Touching the identical comparison is a correct response.

In this two-sample DMTS procedure, both sample stimuli have an equal probability of appearing as the correct comparison. Thus, during the sample observation period, the participant cannot predict which one of the sample stimuli will be the correct comparison. High accuracy (>90% correct) indicates no overselectivity with two sample stimuli. At the other extreme, accuracy at or near chance levels (33%) indicates an overall failure to perform the matching task. Chance-level accuracy scores provide no information relevant to the evaluation of overselectivity. Intermediate accuracy scores (e.g., approximately 67%) indicate overselectivity. The participant was able to match only one of the two sample stimuli, as follows: On those trials in which that stimulus appeared as a comparison (half of the trials in the session), the participant was always correct. On the remaining trials, performance was at chance levels. The intermediate accuracy score for the entire session results from averaging scores from both types of trials (for a detailed analysis, see Dube & McIlvane, 1997).

In the standard DMTS task, the sample observation period continues until the participant touches the sample display area. Thus, the response to the samples (touching) is the same for all stimuli. A procedural modification that may improve discrimination is to require *differential observing responses* (DOR), that is, a different response for each sample stimulus (e.g., Cohen, Brady, & Lowry, 1981; Urcuioli & Callender, 1989). For example, Constantine and Sidman (1975) found that DMTS accuracy with pictures improved in young men with severe mental retardation when they were required to name the sample pictures (cf. Geren, Stromer, & Mackay, 1997). Gutowski, Geren, Stromer, and Mackay (1995) showed that stimulus overselectivity was reduced in 2 individuals with moderate mental retardation when they were required to name the stimuli (pictures of common objects) in a two-sample DMTS task.

Naming, however, is not always available as a differential observing response. Some students in special education classes may be unable to produce spoken or gestural names, and naming is not an option with unfamiliar stimuli. The present study examined a generalized, nonverbal differential observing response procedure. This procedure prompted participants to make simultaneous identity-matching responses during the sample observation period.

METHOD

Participants

Three individuals with mental retardation participated. Pseudonyms, ages, standardized test scores, and diagnostic information from records are shown in Table 1. Prior to this experiment, all had been (a) assessed for snack-food preferences, (b) trained to exchange plastic poker-chip tokens for foods, and (c) given an assessment of matching-to-sample skills (Dube, Iennaco, & McIlvane, 1993). Dawn and Ellen, who could perform generalized identity matching with high accuracy, were next given additional matching-to-sample pretests for this study. Bob, who was unable to perform accurate identity matching, was first trained to do so (with

Table 1
Participant Characteristics

Participant	CA[a]	PPVT[b]	EOWP VT[c]	Diagnostic description[d]
Bob	15	2.4	2.9	PDD[e]
Dawn	19	6.4	7.7	PDD, moderate MR[f]
Ellen	13	2.0	2.0	MR

[a] Chronological age.
[b] Peabody Picture Vocabulary Test–Revised age-equivalent score.
[c] Gardner Expressive One-Word Picture Vocabulary Test (Revised) age-equivalent score.
[d] From students' records.
[e] Pervasive developmental disorder.
[f] Mental retardation.

the sorting-to-matching procedures reported in Serna, Dube, & McIlvane, 1997) and was then given the pretests for this study.

General Procedure

Experimental sessions of 10- to 15-min duration were conducted 3 or 4 days per week in a quiet area at the participants' schools. Participants sat before a Macintosh Plus® computer with a 9-in. black-and-white video display and a touch-sensitive screen. The computer controlled stimulus presentation, response recording, and data collection.

The stimuli were black, nonrepresentative forms (see Figure 1 for examples), approximately 1 cm by 1.5 cm, displayed on a white

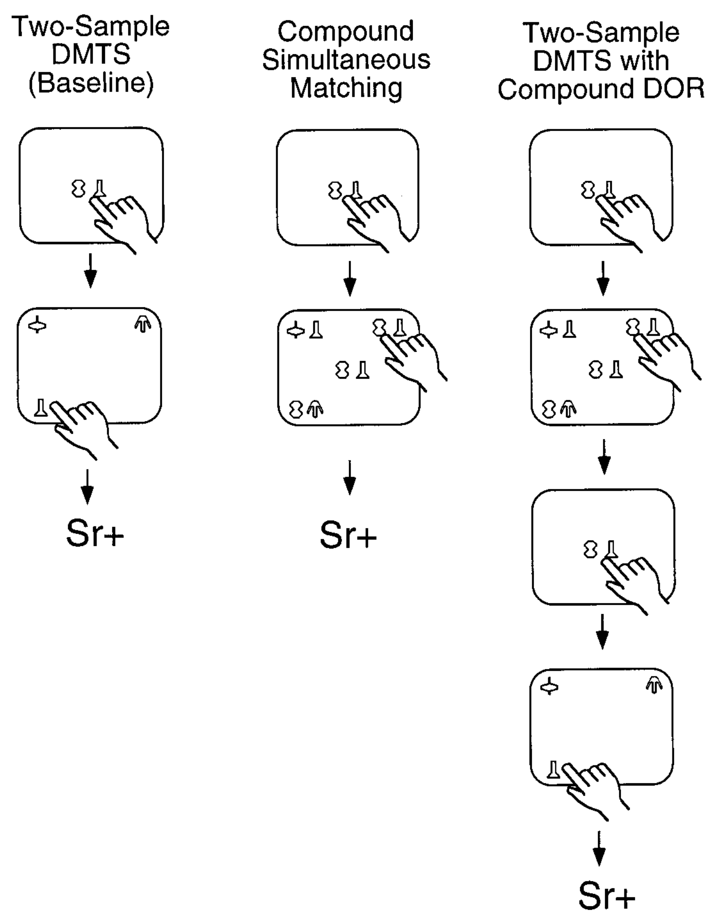

Figure 1. Matching-to-sample procedures. Illustrations show participants making correct responses. Sr+ indicates that the immediately preceding response would be followed by a token reinforcer.

background. Within each session, different stimuli appeared on every trial. The stimuli for each session were drawn at random, without replacement, from a pool of 180 forms.

During 3-s intertrial intervals (ITIs), the display screen was blank. Trials that ended with correct responses were followed by a brief auditory-visual computer display and a token presented by the experimenter. Trials that ended with errors were followed only by the ITI.

Tokens were exchanged for snack foods after sessions. Participants bought small quantities of foods by placing the tokens on a "price tag," a piece of cardboard with 10 token-sized circles. The participant placed a token on each circle, and then the experimenter presented the food and collected the tokens; this process was repeated until no tokens remained (if a price tag was only partially filled for the last purchase, the quantity of food was reduced accordingly).

Matching-to-Sample Pretests

One-sample simultaneous matching to sample. Trials began when a sample stimulus appeared in the center of the screen. When the participant touched the sample, three comparison stimuli appeared in three corners of the screen and the sample remained displayed. One comparison was identical to the sample, and touching it was the correct response. Touching either of the nonidentical comparisons was an error.

One-sample delayed matching to sample (DMTS). Trials were the same as one-sample simultaneous matching, except that the sample stimulus disappeared when the participant touched it. The comparison stimuli appeared immediately (i.e., 0-s delay).

Two-sample simultaneous matching. Trials were similar to one-sample simultaneous matching, except that two sample stimuli were presented on each trial. The samples were displayed side by side, 1.75 cm center to center. The comparisons were single stimuli, and the correct comparison was identical to one of the samples. Over trials, stimuli in the left and right sample positions were correct equally often.

Two-sample DMTS. Trials were the same as two-sample simultaneous matching, except that the sample stimuli disappeared when the participant touched them. (The procedure is illustrated in the left column of Figure 1.) The delay interval from sample offset to comparison onset was 0 s.

Compound simultaneous matching pretests. Trials were the same as two-sample simultaneous matching, except that the comparison display consisted of three pairs of stimuli. (The procedure is illustrated in the center column of Figure 1.) The correct pair of comparisons matched the sample stimuli exactly. Each of the two pairs of incorrect comparisons had one stimulus that matched one of the samples and one that did not match. The matching stimulus was on the left in one pair of incorrect comparisons and on the right in the other pair.

Pretest outcomes. Every pretest session consisted of 36 matching-to-sample trials. Participants were given successive sessions of each type of pretest until performance was stable. The criterion for stability was three consecutive sessions with accuracy scores greater than 50%, in which individual sessions scores did not deviate from the three-session mean by more than 10%. There was one exception: Ellen was given only one two-sample simultaneous matching test because of a scheduling error. Table 2 shows mean accuracy scores for the last three sessions of each type of pretest (given Ellen's exception). In every case in which the number of sessions is greater than three, accuracy scores were initially lower and improved with practice. As Table 2 shows, all participants had scores of at least 98% for both one-sample tasks, at least 94% for the two-sample simultaneous matching task, and in-

Table 2
Pretest Accuracy Scores (Percentage Correct)

Matching task	Participants Bob	Dawn	Ellen
One-sample simultaneous	99 (3)	98 (3)	99 (5)
One-sample delayed	98 (3)	100 (3)	98 (3)
Two-sample simultaneous	96 (3)	94 (3)	97 (1)
Two-sample delayed	69 (3)	66 (9)	71 (3)
Compound simultaneous	97 (3)	91 (12)	99 (7)

Note. Each score is the mean for the final three sessions for each type of matching task (except for Ellen's two-sample simultaneous matching task). Numbers in parentheses show the total number of test sessions. See text for details.

termediate accuracy scores ranging from 66% to 71% for the delayed two-sample task.

All participants also had high accuracy scores for compound simultaneous matching pretests. When initially low scores for Dawn did not improve with practice within three sessions, she was given sessions that began with the two-sample simultaneous matching task and then switched to the compound simultaneous matching task at some point in the session. Initially, sessions began with 18 two-sample trials, and this number was reduced to zero over six sessions.

Immediately after the pretests, participants were adapted to intermittent reinforcement for compound simultaneous matching. They were given sessions consisting of compound simultaneous matching trials only (as in the pretest), and the reinforcement schedule was gradually thinned until only one response in six (on average) was followed by differential consequences (variable-ratio [VR] 6).

Experimental Conditions

The investigation was designed to evaluate the effects of a DOR procedure on two-sample DMTS. An initial baseline condition measured DMTS accuracy. Then, the DOR procedure (described below) was evaluated, followed by a return to the baseline condition.

Baseline. Baseline conditions consisted of six sessions of two-sample DMTS. There were 36 trials in each session.

Compound DOR. The compound DOR procedure embedded a compound simultaneous matching trial within the sample-observation period of the DMTS trial, as shown in the right column of Figure 1. The first touch to the sample display area produced a comparison display consisting of three pairs of stimuli, and the samples remained displayed. As in the compound simultaneous matching pretest, the correct pair of comparisons both matched the sample stimuli, and each of the two pairs of incorrect comparisons had only one stimulus that matched the samples (cf. Allen & Fuqua, 1985; Schreibman, Charlop, & Koegel, 1982). Thus, consistently correct responding verified discrimination of both sample stimuli; comparison selections that were based on only one of the samples would be correct on only half of the trials on average.

On each trial, both the correct and the incorrect comparison stimuli for the final DMTS response were stimuli that had been presented as comparisons during the DOR. In this way, the participant's final response could not be based on relative novelty, because all of the comparison stimuli had been presented earlier in that trial.

In contrast to the procedure during compound simultaneous matching pretests, differential observing responses on DMTS trials were never followed by differential reinforcement or any kind of informative feedback. The trial continued as shown in Figure 1 regardless of which comparison pair was selected. In this way, the comparison stimuli presented at the end of the trial had no immediate reinforcement histories and thus no previously established positive or negative stimulus functions. For clarity and convenience, the task will be referred to as *compound simultaneous matching* when selections of comparison stimuli produced differential

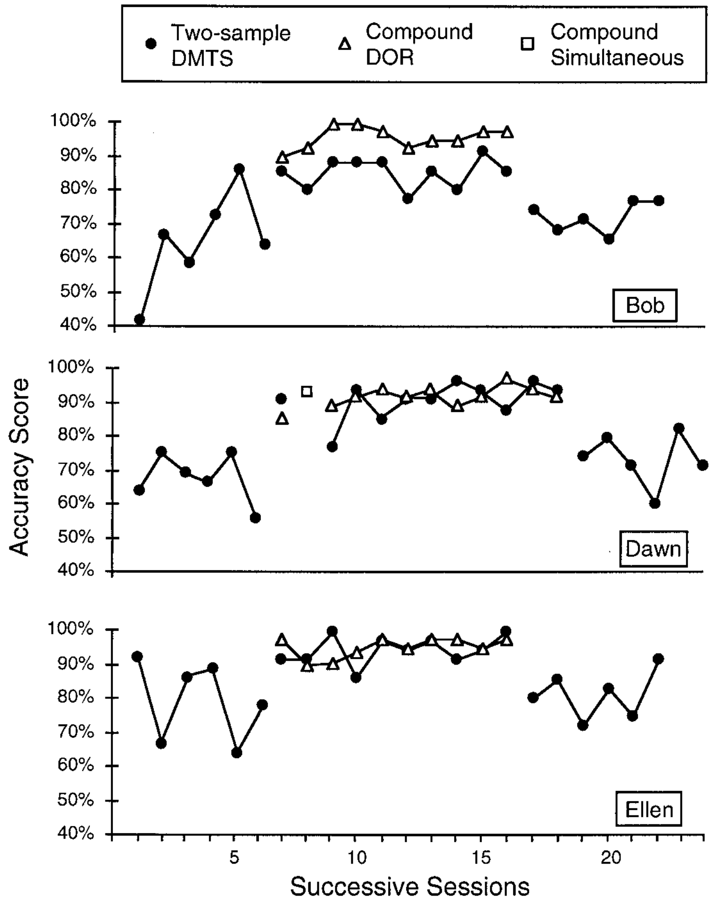

Figure 2. Individual-session accuracy scores. Sessions with one point are baseline DMTS sessions with no DOR procedures (the first six sessions and last six sessions for each participant). Sessions with two points are from the compound DOR condition. The open triangles show accuracy for the DOR portions of the trials (the intermediate response in the right column of Figure 1), and the filled points show accuracy for the final DMTS response. The open square in Dawn's plot shows accuracy for one session of compound simultaneous matching.

consequences (e.g., in the pretests) and as *compound DOR* when the task was embedded in a DMTS trial and never produced differential consequences.

So that responding to DOR displays would not be extinguished because of nonreinforcement, sessions included a small number of compound simultaneous matching trials that ended with differential consequences. Sessions in the compound DOR condition consisted of 42 trials: 36 DMTS trials with the compound DOR plus six compound simultaneous matching trials. The compound DOR condition continued for 10 sessions, with one exception, explained with the results.

RESULTS

Session-by-session accuracy scores for individual participants are shown in Figure 2. Ellen's baseline scores were slightly higher in the first baseline condition (M = 79%) than in her two-sample DMTS pretest (71%). This increase in accuracy may have resulted from her recent history with the compound simultaneous matching procedure during the final pretest and the subsequent sessions that introduced intermittent reinforcement. Accurate compound simultaneous matching required observing both members of pairs of stimuli. In the baseline condition that followed, she apparently continued to observe both stimuli during the sample observation period on some, but not all, of the two-sample DMTS trials.

In the first session with the compound DOR procedure, accuracy on the DOR portions of the trials was at least 90% for Bob and Ellen and 86% for Dawn. High accuracy was expected because of the compound simultaneous matching pretest results. Dawn's accuracy on the compound DOR portions of the trials, however, indicated that the procedure did not set the occasion for reliable discrimination of both sample stimuli. To recover her compound matching performance, she was given one review session consisting of 36 compound simultaneous matching trials with the VR 6 reinforcement schedule. Her accuracy was 94% in this session (open square point in Figure 2), and the DOR procedure was resumed in the following session.

Figure 2 shows that the compound DOR procedure produced clear and immediate increases in DMTS accuracy for Dawn and Ellen and a modest increase in accuracy for Bob. In the baseline condition that followed the compound DOR condition, accuracy returned to previous intermediate levels for all participants.

DISCUSSION

The results show that stimulus overselectivity can be greatly reduced with a DOR procedure that controls observing behavior and sets the occasion for discrimination of all stimuli. Although overselectivity was greatly reduced by requiring DORs, the improvement did not continue after the procedure was discontinued. That is, mere exposure to the DOR procedure was not sufficient to eliminate overselectivity when observing responses were no longer prompted. One important goal for further research is to find ways to teach students to observe multiple stimuli when there are no DOR requirements in the same way that they do when there are such requirements. In general, behavior analysts are challenged to understand why certain individuals do not display or maintain behavior of which they are manifestly capable when doing so would increase the frequency of reinforcers obtained (cf. Constantine & Sidman, 1975). One possibility is that the individual is not sensitive to the difference between the lower and higher reinforcement frequencies. A related possibility concerns our token exchange procedure: Participants accumulated tokens throughout the session and then exchanged them all after the session. For example, a participant had 24 tokens after a baseline session with 67% accuracy and 39 tokens after a DOR session with 92% accuracy on DMTS trials (33 of 36 on DOR-DMTS trials and 6 of 6 on compound simultaneous matching trials); 24 versus 39 tokens could have been a difficult discrimination for some of our participants. One possibility for further study of reinforcement variables is to investigate within-session token exchanges whenever a fixed number have been earned.

As we noted earlier, this paper is intended to be a bridge study. We used arbitrary forms as stimuli to minimize the influence of preexperimental experience. Previous experimentation has shown, for example, that prior reinforcement histories can determine which specific stimuli gain control in cases of overselectivity (Dube & McIlvane, 1997). Also, the use of unfamiliar stimuli reduced the possibility that a participant might produce names for the stimuli; if he or she routinely named the stimuli, the task would no longer be strictly a measure of visual stimulus control.

Matters of experimental control notwithstanding, it is appropriate to ask how our findings might influence practices in applied settings. Our results show clearly that nonverbal DOR procedures that capitalize on generalized identity-matching repertoires can be effective in reducing stimulus overselectivity. It seems reasonable to anticipate that procedures modeled after those we have described would prove useful with everyday academic stimuli. For example, consider training to teach matching relations between pictures of clothing items and the corresponding printed words. In cases in which sample stimuli are pictures, nonverbal compound DORs like those described in this paper could help to eliminate overselective stimulus control by isolated features of the pictures (e.g., for the DOR, a sample picture of brown loafers, with comparison pictures of brown loafers, black loafers, and brown oxfords; cf. Burke & Cerniglia, 1990; Rosenblatt, Bloom, & Koegel, 1995; Schreibman, 1997). When the sample stimuli are printed words, nonverbal DORs could be programmed by inserting printed-word identity-matching trials in which the incorrect comparison stimuli have letters in common with the sample (e.g., sample HAT, with comparison stimuli HAT, CAT, HOT, and HAD). Another option for nonverbal DORs with printed words is constructed-response matching to sample, in which the letters that make up the word are matched individually and sequentially (Dube, McDonald, McIlvane, & Mackay, 1991; Mackay, 1985). All of these procedures can be implemented either on the computer or the tabletop, and they may find wide application with individuals who might otherwise respond overselectively.

REFERENCES

Allen, K. D., & Fuqua, R. W. (1985). Eliminating selective stimulus control: A comparison of two procedures for teaching mentally retarded children to respond to compound stimuli. *Journal of Experimental Child Psychology, 39,* 55–71.

Bickel, W. K., Richmond, G., Bell, J., & Brown, K. (1986). A microanalysis of the controlling stimulus-response relations engendered during the assessment of stimulus overselectivity. *The Psychological Record, 36,* 225–238.

Birnie-Selwyn, B., & Guerin, B. (1997). Teaching children to spell: Decreasing consonant cluster errors by eliminating selective stimulus control. *Journal of Applied Behavior Analysis, 30,* 69–91.

Burke, J. C., & Cerniglia, L. (1990). Stimulus complexity and autistic children's responsivity: Assessing and training a pivotal behavior. *Journal of Autism and Developmental Disorders, 20,* 233–253.

Cohen, L. R., Brady, J., & Lowry, M. (1981). The role of differential responding in matching-to-sample and delayed matching performance. In M. L. Commons & J. A. Nevin (Eds.), *Quantitative analysis of behavior: Vol. 1. Discriminative properties of reinforcement schedules* (pp. 345–364). Cambridge, MA: Ballinger.

Constantine, B., & Sidman, M. (1975). The role of naming in delayed matching to sample. *American Journal of Mental Deficiency, 79,* 680–689.

de Rose, J. C., de Souza, D. G., & Hanna, E. S. (1996). Teaching reading and spelling: Exclusion and stimulus equivalence. *Journal of Applied Behavior Analysis, 29,* 451–469.

Dube, W. V., Iennaco, F. M., & McIlvane, W. J. (1993). Generalized identity matching to sample of two-dimensional forms in individuals with intellectual disabilities. *Research in Developmental Disabilities, 14,* 457–477.

Dube, W. V., McDonald, S. J., McIlvane, W. J., & Mackay, H. A. (1991). Constructed-response matching to sample and spelling instruction. *Journal of Applied Behavior Analysis, 24,* 305–317.

Dube, W. V., & McIlvane, W. J. (1997). Reinforcer frequency and restricted stimulus control. *Journal of the Experimental Analysis of Behavior, 68,* 303–316.

Geren, M. A., Stromer, R., & Mackay, H. A. (1997). Picture naming, matching to sample, and head injury: A stimulus control analysis. *Journal of Applied Behavior Analysis, 30,* 339–342.

Gutowski, S. J., Geren, M., Stromer, R., & Mackay, H. A. (1995). Restricted stimulus control in delayed matching to complex samples: A preliminary analysis of the role of naming. *Experimental Analysis of Human Behavior Bulletin, 13,* 18–24.

Lovaas, O. I., Koegel, R. L., & Schreibman, L. (1979). Stimulus overselectivity in autism: A review of research. *Psychological Bulletin, 86,* 1236–1254.

Mackay, H. A. (1985). Stimulus equivalence in rudimentary reading and spelling. *Analysis and Intervention in Developmental Disabilities, 5,* 373–387.

Rosenblatt, J., Bloom, P., & Koegel, R. L. (1995). Overselective responding. In R. L. Koegel & L.

K. Koegel (Eds.), *Teaching children with autism* (pp. 33–42). Baltimore: Brookes.

Schreibman, L. (1997). The study of stimulus control in autism. In D. M. Baer & E. M. Pinkston (Eds.), *Environment and behavior* (pp. 203–209). Boulder, CO: Westview.

Schreibman, L., Charlop, M. H., & Koegel, R. L. (1982). Teaching autistic children to use extra-stimulus prompts. *Journal of Experimental Child Psychology, 33,* 475–491.

Serna, R. W., Dube, W. V., & McIlvane, W. J. (1997). Assessing same/different judgments in individuals with severe intellectual disabilities: A status report. *Research in Developmental Disabilities, 18,* 343–368.

Stromer, R., Mackay, H. A., & Stoddard, L. T. (1992). Classroom applications of stimulus equivalence technology. *Journal of Behavioral Education, 2,* 225–256.

Stromer, R., McIlvane, W. J., Dube, W. V., & Mackay, H. A. (1993). Assessing control by elements of complex stimuli in delayed matching to sample. *Journal of the Experimental Analysis of Behavior, 59,* 83–102.

Urcuioli, P. J., & Callender, J. (1989). Attentional enhancement in matching-to-sample: Facilitation in matching acquisition by sample-discrimination training. *Animal Learning & Behavior, 17,* 361–367.

Wacker, D. P. (1996). Behavior analysis research in *JABA*: A need for studies that bridge basic and applied research. *Experimental Analysis of Human Behavior Bulletin, 14,* 11–14.

Received February 23, 1998
Initial editorial decision April 21, 1998
Final acceptance August 24, 1998
Action Editor, Joseph E. Spradlin

STUDY QUESTIONS

1. What is stimulus overselectivity? Give an example illustrating this problem.

2. Describe the procedures involved in a two-sample delayed matching-to-sample (DTMS) task. Why are intermediate accuracy scores (67%) on such a task suggestive of stimulus overselectivity?

3. What was the purpose of the current study?

4. What results from the pretests indicated that participants' discriminative responding was not always overselective?

5. What is an observing response, what is its function, and what was the specific observing response taught for the two-sample DMTS task?

6. Summarize the results obtained during the experimental conditions. In what sense was the differential observing response (DOR) generalized and in what sense was it not generalized?

7. What feature of the reinforcement contingencies may have affected the results? Can you suggest an alternative?

8. Provide an example of how the techniques described in this study might be applied to teaching situations with individuals with severe disabilities.

Study questions prepared by Gregory P. Hanley and Eileen M. Roscoe, The University of Florida